Degree of Confounding Bias Related to Smoking, Ethnic Group, and Socioeconomic Status in Estimates of the Associations Between Occupation and Cancer

Jack Siemiatycki, PhD; Sholom Wacholder, PhD; Ronald Dewar, MSc; Elizabeth Cardis, PhD; Celia Greenwood, MSc; and Lesley Richardson, MSc

In occupational cancer epidemiology, many studies are carried out without access to information on smoking and other potential confounding variables. It is unclear whether such deficiencies are likely to cause serious bias in estimates of cancer-occupation associations. An empiric investigation was carried out to determine the effect of inclusion or exclusion of three variables—smoking, ethnic group, and socioeconomic status—on estimates of odds ratios (OR) between 25 occupations and three types of cancer—lung, bladder, and stomach. Of the 75 associations studied, only one OR was distorted by more than 40% when comparing unadjusted with adjusted estimates; three were distorted by between 30% and 40%; four others by between 20% and 30%. Of the eight associations which were distorted by more than 20%, seven involved lung cancer and one involved bladder cancer; none involved stomach cancer. An additional analysis was carried out on the 25 lung cancer-occupation associations to determine whether the nature of the stratification on smoking (ie, whether crude or "precise" categories were used) gave different OR estimates. The differences in ORs induced by different parametrizations of the smoking variable were relatively small.

Our results support the view that relative risks between lung cancer and occupation in excess of 1.4 are unlikely to be artifacts due to uncontrolled confounding. For bladder and stomach cancer, the corresponding cut point may be as low as 1.2. In studies of occupation and cancer, uncontrolled confounding due to smoking and social class may not be as serious a threat to the integrity of results as is sometimes feared.

Much of the activity of epidemiologic research revolves around the thorny issue of controlling for confounding factors. There has been considerable theoretical development in the past few decades concerning the mechanisms of confounding and statistical methods to adjust for confounding factors.[1–4] Different authors have given varying, and even conflicting definitions of confounding.[5] A covariate can be said to confound the association between an exposure and a disease if the measure of association between exposure and disease differs according to whether or not the estimate is adjusted for the covariate. For a covariate to be a confounder, it must be associated with the exposure of interest (in the nondiseased base population) and with the disease under study (in the base population free of the exposure of interest).

If it is not possible to control for confounding in the design of a study, the best estimate of an association is obtained when the true confounding covariates are known, have been measured precisely, and have been adjusted for in the analysis. Adjusting for more covariates than is necessary will not lead to bias but it

From the Epidemiology and Preventive Medicine Research Centre, Institut Armand-Frappier, Laval-des-Rapides, Quebec, Canada H7V 1B7 (Dr Siemiatycki, Mr Dewar, Dr Cardis, Ms Richardson); and the Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada H3A 1A2 (Dr Siemiatycki, Dr Wacholder, Mr Dewar, Ms Greenwood). Dr Wacholder is currently at the National Cancer Institute, Bethesda, MD 20829. Ms Greenwood is currently at the Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5.

Address reprint requests to Dr J. Siemiatycki, Epidemiology and Preventive Medicine Research Centre, Institut Armand-Frappier, 531 des Prairies Blvd, Laval-des-Rapides, Quebec, Canada H7V 1B7.

0096-1736/88/3008-617$02.00/0

can reduce precision.[6] Failure to adjust for true confounders can lead to bias; we refer to this as confounding bias.

The most common study design in occupational cancer epidemiology has been the historic cohort study based on company or union records, in which a work force's mortality experience is compared to that of a standard population, usually the entire country. In these studies adjustments can usually be made for sex, age, race, and calendar year but, typically, information on several other potential confounders (eg, smoking and social class) is not available. Another important study design in this field is exemplified by the United Kingdom Registrar General's Decennial Report,[7] in which death certificate information on cause of death and occupation is used to derive proportionate or standardized mortality ratios. Here again, information on smoking is usually unavailable in such studies. It is reasonable to question whether cancer-occupation associations estimated in such studies have been distorted by confounding.

From theory, it has long been known that the confounding bias in estimation of a disease-exposure relative risk cannot be greater than the weaker of the two associations: confounder-disease and confounder-exposure.[1-4,8] Stronger statements can be made about the upper limits of confounding bias, which indicate that the degree of confounding bias is likely to be much less than the lesser of the two associations mentioned above.[4,9] How this theory translates into practice depends, of course, on the particularities of the confounder-disease and confounder-exposure associations. Although relationships between cancer, occupation, and covariates undoubtedly differ somewhat from place to place and study to study, there may be enough in common that empiric evidence from one situation illustrates the likely order of magnitude of confounding effects.

Asp[10] and Blair et al[11] provided some empiric evidence that the absence of smoking information was not a serious source of confounding in cancer-occupation studies. Since the generalizability of such empiric evidence is questionable, we intended to replicate and extend their findings in a different milieu, using a different approach. We chose to evaluate the possible confounding effect on cancer-occupation associations of three variables which have been shown to be related to cancer risk and which are often associated, in one way or another, with occupation—namely smoking, ethnic group, and socioeconomic status (SES). These are variables which, unlike age, sex, and race, are seldom available for members in historic cohort studies. Confounding bias related to smoking, ethnic group, and SES was assessed in the context of estimating the associations between 25 occupations and three sites of cancer—lung, bladder, and stomach. The battery of 75 associations examined covers a wide range of cancer-occupation-confounder situations and may thus be relevant to other investigators confronting similar situations. Specifically, the present report addresses the following questions concerning the estimation of these associations: (1) Is the estimate affected by whether the covariates (smoking, socioeconomic status and ethnic group) are adjusted for individually, in combination, or not at all? (2) Does it matter whether the smoking information used is crude or detailed?

The data used to address these issues were collected as part of a large case-control study of occupational factors in cancer. Several cancer sites were included. For each subject, information was obtained on the job history and on a variety of social and demographic variables. Thus it was possible to assess cancer-occupation associations with and without controlling for covariates and with different forms of control for covariates.

Methods

Since 1979 we have been conducting a cancer case-control study designed to assess associations between several sites of cancer on the one hand and scores of occupational exposures on the other. Eligibility criteria for cases included the following: male, aged 35 to 70 years, resident in the Metropolitan Montreal area, with a newly diagnosed tumor of any of the 19 sites of cancer selected for study. All major hospitals in the area have participated, thus providing a population-based series. In addition, a general population control series was selected from electoral lists. Each eligible subject was approached for an interview. The interview concerned details of the subject's job history as well as information on potential confounders. For each case series, control subjects could be selected among the other cancer case subjects, among the population control subjects, or among both. Detailed explanations of the design and study methods have been published elsewhere,[12,13] as has a set of substantive results.[14]

The present analyses were based on 4,079 subjects who were interviewed between September 1979 and December 1985. We decided to focus attention on three sites of cancer which are among the most common in our data set and which have distinct relationships to the potential confounders: lung, with 791 cases; bladder, with 461 cases; and stomach, with 231 cases. Separate analyses were carried out for each site of cancer as a case series. The pool of control subjects consisted of a group of 533 "population controls" plus a group of 2,063 patients with cancer of one of the other 16 sites. For the analyses of bladder cancer and stomach cancer, all 2,596 eligible control subjects were used. Lung cancer was exceptional in that we ascertained and interviewed lung cancer case subjects only in 3 of the 6 years of the study. To ensure comparability we restricted the pool of control subjects for lung cancer to those subjects interviewed during the same years in which lung cancer was ascertained. Thus there were 1,305 control subjects for the lung cancer case subjects.

Each job held by a subject was classified into a seven-digit job classification code according to the standard Canadian Dictionary of Occupation Titles.[15] For the present purpose, this was regrouped into 57 categories based on the first two or three digits of the classification

code. A person was considered to be exposed to an occupation if he worked for at least 6 months at that occupation.

The objects of attention are the odds ratios between the three sites of cancer and the various occupation categories. For each cancer-occupation association, several estimates were made of the odds ratio, each estimate adjusted for a different set of covariates. It is the variation in odds ratios across these sets of covariates, for a given association, that is of interest here. The degree of variation indicates the extent to which it matters whether the putative confounders are adjusted for or not.

In fact, the observed variation in estimates of an odds ratio across different sets of covariates has two components: one due to true confounding effects of one or more of the putative confounders and another due to sampling variation. Since we are interested in the component due to confounding, it is desirable to minimize the component due to sampling variation; this was accomplished by restricting attention to those results based on relatively large sample sizes. That is, rather than analyzing the associations between the three sites of cancer and all 57 job categories, we used only the 25 occupations which had the most subjects. The percentage of all 4,079 study subjects who had ever worked in these selected occupations (ie, the percentage "exposed") ranged from 2.6% to 23.3%. Table 1 shows the occupations selected and the numbers "exposed" to each occupation among each of the three case series.

The interview provided information on many potential confounders, including age, cigarette smoking, ethnic group, and socioeconomic status. As stated above, age is usually available in epidemiologic studies and it would be of purely academic interest to determine the confounding effects of its hypothetical absence. Therefore, we routinely included age as a variable to be adjusted for in all analyses. To facilitate communication, we will use the phrase "unadjusted" odds ratio to signify one that has been adjusted for age only; any adjusted odds ratio that we refer to will have been adjusted for age as well as for the other variables mentioned. Sex and race, two variables that are similar to age in that they are usually available, were not at issue in our study, since the data set includes only men and virtually only white subjects.

Although there are a large number of ethnic groups in the Montreal area, the main group consists of French Canadians who constitute about 65% of the population and who have maintained a distinct cultural identity. Most of the rest are also of European origin, the largest subgroup being of British/Irish descent. Socioeconomic

TABLE 1

Occupations Selected for Study, Number of Subjects in Each Occupation, and Numbers of Exposed Cases for Each of the Three Site Series as Well as for Controls

Occupation Group	Total Exposed No.*	Case Subjects			Control Subjects†
		Bladder Cancer	Stomach Cancer	Lung Cancer	
Administrative	533	70	19	62	382
Sciences, engineering	227	30	10	33	154
Teaching	138	25	7	14	92
Clerical	803	91	32	157	523
Sales	934	126	41	151	616
Protective services	827	94	49	176	508
Food services	291	30	21	62	178
Other services	359	32	22	87	218
Farming	377	41	33	81	222
Forestry	172	13	22	48	89
Metal processing	108	11	3	33	61
Food processing	236	25	16	45	150
Metal machining	189	25	7	39	118
Metal shaping	259	35	10	60	154
Metal products fabrication	133	13	6	34	80
Electrical equipment fabrication	137	12	12	29	84
Wood products fabrication	102	13	8	26	55
Textile products fabrication	248	24	18	41	165
Mechanics	375	33	23	78	241
Other product fabrication	198	20	6	51	121
Excavating, paving	141	13	15	35	78
Electrical construction	106	17	7	17	65
Other construction	724	78	52	161	433
Motor transport	651	83	50	144	374
Materials handling	379	31	22	87	239
Total no. of men‡	**4,079**	**461**	**231**	**791**	**2,596**

* The number of subjects who, at any time in their careers, worked for at least 6 months in the occupation. The four subsequent columns present an exhaustive, mutually exclusive subdivision of the "total number exposed." For each occupation group the "unexposed" consist of all other men in the study, including those in occupations other than the 25 listed here. Note that the same man may appear in a few occupation groups, depending on his job history. Thus the sum of the elements in each column far exceeds the number of men in the study (last row).

† Controls comprise men with other types of cancer and population controls.

‡ The total shown here counts each man only once and includes those who worked in the 25 listed occupations and those who worked in other occupations.

status was assessed in the questionnaire by asking the respondent about family income in the preceding year. For those 20% who refused to divulge this information, we imputed socioeconomic status from the mean family income of the census tract of residence. For smokers, information was available on age when smoking began, age when smoking ended (for ex-smokers), and average amount smoked per day. We were thus able to derive the duration of smoking, the intensity (or amount per day), and a cumulative pack-year variable.

There are various possible procedures for estimating the odds ratio (OR) between disease and exposure, adjusting for various covariates. The most frequently used ones are logistic regression and the Mantel-Haenszel estimator.[4] Because of the large number of analyses to be carried out, it was more convenient to use Mantel-Haenszel methods. A special program has been developed for the purpose.[16] The results from logistic regression analysis of the same data are not likely to be very different.

For each of the three types of cancer (lung, bladder, and stomach), an analysis was carried out to determine the variation in odds ratio estimates according to whether or not smoking, ethnic group, and SES were included in the analysis as confounders.

Eight odds ratio estimates were calculated for each of the 3 × 25 site-occupation associations. One was adjusted only for age; the others were adjusted for age plus each of the seven possible combinations of the three covariates: (1) smoking only, (2) ethnicity only, (3) SES only, (4) smoking and ethnicity, (5) smoking and SES, (6) ethnicity and SES, (7) smoking, ethnicity, and SES. In those analyses, ethnicity was dichotomized (French/other), SES was dichotomized (below 30 percentile/above 30 percentile), and smoking was trichotomized according to cumulative amount smoked in cigarette-years (0/1–599/600+).

The variation in odds ratio estimates of the same association describes the possible impact of confounding by these variables. The ratios of adjusted to unadjusted odds ratios were used as indices of confounding bias. Implicit in our approach is the assumption that the degree of confounding bias can be meaningfully estimated independently of the strength of a disease-exposure relationship. That is, the same confounding bias may distort a true odds ratio of 1.0 to become 1.5 or a true odds ratio of 2.0 to become 3.0. Since it is the magnitude of the bias rather than its direction which is of interest, it is convenient to compute all measures of confounding bias with the larger odds ratio in the numerator and the smaller odds ratio in the denominator. Thus,

$$\text{confounding bias ratio} = \text{larger OR}/\text{smaller OR}$$

where the two ORs are estimates of the same disease-occupation odds ratio, but incorporating different confounder stratifications.

Because of the strong association between smoking and lung cancer, it is of interest to determine whether different measures and categorizations of smoking as a confounder variable lead to different results. For this purpose, seven cigarette smoking variables were created:

1. never/ever
2. never/daily intensity 1–19 cigarettes/20+ cigarettes
3. never/daily intensity 1–9 cigarettes/10–19 cigarettes/20–39 cigarettes/40+ cigarettes
4. never/duration 1–29 years/30+ years
5. never/duration 1–19 years/20–29 years/30–49 years/50+ years
6. never/cumulative amount 1–199 cigarette-years/200–599 cigarette-years/600–1,499 cigarette-years/1,500+ cigarette-years.

The cut points were chosen to maximize differences in risk among subcategories and to provide approximately similar frequency distributions across the three scales of duration, intensity, and cumulative amount. To determine the effect of different types of control for cigarette smoking, we estimated each lung cancer-occupation odds ratio eight times, once without any control for smoking and once with each of the seven smoking variables as confounders.

Results

Table 2 shows percentages of heavy smokers, low SES, and French Canadians among the 25 occupations used for these analyses. The occupations have varying pro-

TABLE 2
Selected Characteristics of the 25 Occupation Groups Under Study, Among the 2,596 Subjects Who Were Used as Controls*

Occupation Group	Total Exposed, No.	Heavy Smokers, %	French, %	Low Income, %
Administrative	382	51.8	50.8	14.1
Sciences, engineering	154	47.4	35.1	10.4
Teaching	92	32.6	55.4	17.4
Clerical	523	57.6	60.0	25.3
Sales	616	59.1	61.4	23.9
Protective services	508	67.9	55.5	30.3
Food services	178	64.6	68.0	41.0
Other services	218	67.9	62.4	47.7
Other farming	222	57.2	56.8	35.6
Forestry	89	76.4	86.5	43.8
Metal processing	61	59.0	52.5	44.3
Food processing	150	56.7	61.3	38.7
Metal machining	118	67.0	56.8	28.0
Metal shaping	154	61.7	61.0	39.0
Metal fabrication	80	68.8	72.5	36.2
Electrical fabrication	84	52.4	52.4	29.8
Wood fabrication	55	61.8	63.6	43.6
Textile fabrication	165	58.2	47.9	29.7
Mechanics	241	67.2	60.6	29.9
Other fabrication	121	67.8	62.8	32.2
Excavating	78	64.1	55.1	48.7
Electrical construction	65	61.5	63.1	21.5
Other construction	433	65.1	65.8	34.0
Motor transport	374	69.2	73.5	39.0
Materials handling	239	66.5	69.0	38.1
Among all controls	**2,596**	**58.2**	**59.9**	**29.9**

* This table concerns the subjects used as controls in the case-control analyses, namely, the population controls and cancer patients used as controls.

files for these variables. For instance, among teachers there were 32% heavy smokers, whereas the corresponding figure among motor transport workers was 69%. Among scientists and engineers, 35% were French, whereas the corresponding figure among forestry workers was 86%. Table 3 addresses the second side of the confounding triangle, the covariate-disease association. Lung cancer is associated with each of the three covariates, most strongly, as expected, with cigarette smoking. Both bladder and stomach cancer appear to be associated with cigarette smoking, and stomach cancer is associated with SES. Since the covariates are associated with both the diseases and the occupations, there is opportunity for confounding.

Table 4 shows the odds ratios across eight covariate combinations between lung cancer and each of the 25 selected occupations. The table also presents, for each lung cancer-occupation association, three measures of confounding bias: one based solely on the ratio between the smoking adjusted odds ratio vis-à-vis the unadjusted odds ratio, a second based on the most adjusted compared to the unadjusted odds ratio, and another based on the maximum discrepancy between the unadjusted OR and any of the adjusted ORs. Nearly all values of the first confounding bias factor were below 1.20. The only exception was that for teachers, for whom it was 1.62. This is in part a reflection of the fact, as shown in Table 2, that teachers had a particularly deviant smoking profile. The third measure of confounding bias, based on the maximum discrepancy from the unadjusted odds ratio is of course the one that produces the greatest values and represents a "worst-case" indicator of bias if confounders are not adjusted for. Among the 25 occupations under consideration, five have maximum confounding bias factors above 1.20: teachers (1.72), forestry workers (1.38), scientists and engineers (1.35), electrical workers in construction (1.29), and other service workers (1.25). It is noteworthy that the most

TABLE 3
Odds Ratios (OR) Between Each of the Three Case Series and Each of the Covariates

Disease	Covariate Categories Compared	OR*
Lung cancer	SES†: low *v* other	1.5
	Ethnicity: French *v* other	1.7
	Smoking: light *v* never	5.0
	Smoking: heavy *v* never	15.8
Bladder cancer	SES: low *v* other	1.2
	Ethnicity: French *v* other	1.0
	Smoking: light *v* never	1.5
	Smoking: heavy *v* never	2.7
Stomach cancer	SES: low *v* other	1.4
	Ethnicity: French *v* other	1.0
	Smoking: light *v* never	1.3
	Smoking: heavy *v* never	1.5

* Mantel-Haenszel odds ratio controlling for age.
† SES, socioeconomic status.

TABLE 4
Odds Ratios (OR) Between Lung Cancer and Each of 25 Occupation Categories, Adjusting for Eight Different Combinations of Smoking, Socioeconomic Status (SES) and Ethnic Group

Occupation	Odds ratios with the following covariates included in addition to age								Confounding bias factor*			Most Discrepant OR_i†
	None OR_0	Smoking OR_1	SES OR_2	Ethnic OR_3	Smoking/ SES OR_4	Smoking/ Ethnic OR_5	Ses/ Ethnic OR_6	Smoking/ Ethnic/ SES OR_7	Smoking I	Smoking/ Ethnic/ SES II	Maximum Discrepancy III	
Administrative	0.48	0.51	0.53	0.51	0.56	0.54	0.55	0.58	1.07	1.21	1.21	OR_7
Sciences, engineering	0.65	0.76	0.73	0.71	0.84	0.81	0.79	0.88	1.17	1.35	1.35	OR_7
Teaching	0.44	0.71	0.48	0.46	0.76	0.70	0.50	0.74	1.62	1.70	1.72	OR_4
Clerical	1.01	1.02	1.03	1.01	1.04	1.02	1.03	1.04	1.00	1.02	1.02	OR_7
Sales	0.75	0.72	0.78	0.74	0.74	0.71	0.76	0.73	1.05	1.02	1.05	OR_2
Protective services	1.06	0.95	1.07	1.10	0.95	0.97	1.09	0.96	1.12	1.10	1.12	OR_1
Food services	1.01	0.93	0.94	0.98	0.88	0.92	0.93	0.89	1.09	1.13	1.14	OR_4
Other services	1.30	1.15	1.18	1.33	1.06	1.18	1.23	1.10	1.13	1.18	1.23	OR_4
Farming	1.11	1.16	1.08	1.14	1.14	1.17	1.11	1.16	1.04	1.04	1.05	OR_5
Forestry	1.99	1.70	1.79	1.77	1.56	1.56	1.62	1.45	1.17	1.38	1.38	OR_7
Metal processing	1.68	1.53	1.55	1.76	1.45	1.56	1.63	1.48	1.10	1.13	1.16	OR_4
Food processing	0.90	0.88	0.84	0.89	0.84	0.87	0.84	0.84	1.01	1.06	1.07	OR_2
Metal machining	1.19	1.07	1.18	1.21	1.06	1.08	1.21	1.10	1.12	1.08	1.12	OR_4
Metal shaping	1.22	1.17	1.19	1.24	1.13	1.16	1.22	1.13	1.05	1.08	1.08	OR_7
Metal fabrication	1.30	1.14	1.27	1.26	1.13	1.14	1.23	1.12	1.14	1.16	1.16	OR_7
Electrical fabrication	1.15	1.32	1.15	1.19	1.31	1.34	1.17	1.32	1.15	1.15	1.17	OR_5
Wood fabrication	1.39	1.24	1.29	1.32	1.17	1.20	1.24	1.14	1.11	1.22	1.22	OR_7
Textile fabrication	0.76	0.66	0.78	0.80	0.68	0.68	0.81	0.69	1.15	1.10	1.15	OR_1
Mechanics	1.18	1.08	1.16	1.19	1.07	1.10	1.17	1.09	1.09	1.08	1.10	OR_4
Other fabrication	1.34	1.26	1.33	1.34	1.26	1.26	1.35	1.25	1.06	1.07	1.07	OR_7
Excavating	1.46	1.42	1.37	1.46	1.36	1.41	1.38	1.37	1.03	1.07	1.07	OR_7
Electrical construction	1.15	1.03	1.28	1.09	1.12	1.00	1.22	1.08	1.11	1.06	1.15	OR_5
Other construction	1.27	1.16	1.23	1.25	1.13	1.16	1.22	1.13	1.10	1.12	1.12	OR_7
Motor transport	1.37	1.15	1.25	1.27	1.07	1.10	1.18	1.03	1.19	1.33	1.33	OR_7
Materials handling	1.21	1.16	1.15	1.12	1.11	1.10	1.08	1.07	1.04	1.14	1.14	OR_7

* I = the greater of OR_1/OR_0 or OR_0/OR_1; II = the greater of OR_7/OR_0 or OR_0/OR_7; III = the greater of OR_i/OR_0 or OR_0/OR_i where OR_i is the most discrepant of the seven ORs from OR_0.
† This is the OR_i which is most discrepant from OR_0 and which served to compute factor III.

discrepant of the seven adjusted odds ratio estimates was most often (12 out of 25 times) the most adjusted estimate, OR_7, indicating that the confounding effects of the three confounders were often in the same direction.

Table 5 presents the distribution of the confounding bias factors for each set of covariates represented by a column in Table 4, as well as the distribution of bias factors for the maximum discrepancy between the unadjusted and any of the adjusted estimates. It is clear from this table that the inclusion of smoking as a stratification variable had a greater effect in modifying unadjusted values than did the inclusion of SES or ethnicity. The inclusion of all three covariates together in the analysis produced the greatest effects.

Another set of analyses was carried out on the same 25 associations between lung cancer and selected occupations, this time with varying degrees of detail on smoking as the stratification variable for the Mantel-Haenszel odds ratio estimates. Seven different smoking variables were created ranging from the crudest (never/ever) to a five-category variable integrating information on duration and intensity of smoking. Rather than showing the odds ratios for each occupation and each confounder stratification, we show in Table 6 in summary form the distribution of eight indices of confounding bias. These are analogous to the eight indices shown in Table 5, with the first seven corresponding to the comparison of each form of smoking adjustment with the unadjusted estimate and the last representing a worst-case situation based on the maximum discrepancies between adjusted and unadjusted estimates. The last column shows that five occupations had a confounding bias factor higher than 1.20 with one or other of the seven smoking adjustments: teachers (1.62), scientists and engineers (1.37), forestry workers (1.26), electrical workers in construction (1.24), and motor transport workers (1.24).

The estimates based on the crudest smoking adjustment ("never/ever" status) hardly different from the unadjusted estimates; 20 of 25 were within 10% of the unadjusted value and only one of 25 differed by more than 20%. The three-category smoking stratification variables, especially those based on duration and cumulative cigarette-years, provided greater dispersion from the unadjusted value. The five-category variables provided somewhat greater dispersion than the three category variables.

An analysis was carried out for bladder cancer-occupation associations and another for stomach cancer-

TABLE 5

Distribution of Confounding Bias Factors for Each of Seven Stratifications by Smoking, Socioeconomic Status (SES), and Ethnicity Compared with the Estimate Unadjusted for Those Variables, for the Associations Between Lung Cancer and 25 Occupations*

Confounding Bias Factor	Confounder Stratification†							Based on Maximum Discrepancy Between OR_0 and the OR_is
	Smoking	SES	Ethnicity	Smoking/ SES	Smoking/ Ethnicity	Ethnicity/ SES	All	
1.00–1.10	13	20	23	9	12	18	10	7
1.10–1.20	11	5	2	11	9	5	9	11
1.20–1.30	0	0	0	4	3	2	2	3
1.30–1.40	0	0	0	0	0	0	3	3
1.40–1.50	0	0	0	0	0	0	0	0
1.50–1.60	0	0	0	0	0	0	0	0
1.60–1.72	1	0	0	1	1	0	1	1
Total associations	**25**	**25**	**25**	**25**	**25**	**25**	**25**	**25**

* For each cancer-occupation association, the confounding factor equals the greater of OR_i/OR_0 or OR_0/OR_i, where OR_0 is adjusted for age only, and OR_i is one of the seven estimates adjusted for some combination of smoking, ethnicity, and SES in addition to age.

† The number of strata in each stratification equals the product of the numbers of categories of each covariate included. Age has two categories and was included in all stratifications. Smoking, SES, and ethnicity had three, two, and two categories, respectively. Thus, for example, the column headed "Smoking/SES" had 2 (age) × 3 (smoking) × 2 (SES) = 12 strata in the analyses.

TABLE 6

Distribution of Confounding Bias Factors for Each of Various Stratifications of Cigarette Smoking Compared with the Estimate Unadjusted for Smoking for the Associations Between Lung Cancer and 25 Occupations*

Confounding Bias Factor	Smoking Stratification (No. of Categories Including Nonsmoker)†							
	Ever(2)	Intensity(3)	Intensity(5)	Duration(3)	Duration(5)	Pack-years(3)	Pack-years(5)	Based on Maximum Discrepancy Between OR_0 and the OR_is
1.00–1.10	20	19	15	15	12	13	10	7
1.10–1.20	4	5	9	6	9	11	11	13
1.20–1.30	1	0	0	3	2	0	3	3
1.30–1.40	0	1	1	0	1	0	0	1
1.40–1.50	0	0	0	0	0	0	0	0
1.50–1.62	0	0	0	1	1	1	1	1
Total associations	**25**	**25**	**25**	**25**	**25**	**25**	**25**	**25**

* For each cancer-occupation association, the confounding factor equals the greater of OR_i/OR_0 or OR_0/OR_i, where OR_0 is adjusted for age only and OR_i is one of the seven estimates adjusted for smoking history, in addition to age.

† Each OR was adjusted for age (two categories) in addition to smoking. The numbers of strata were thus twice the numbers indicated in the column headings.

occupation associations, similar to that described in Tables 4 and 5 for lung cancer. By contrast with the lung cancer results, the confounding bias factors for both bladder and stomach cancer were concentrated in the range 1.00 to 1.10. There was little variation across the various combinations of confounders. For illustrative purposes, we show in Table 7 the distribution of bias factor corresponding to two of the confounder combinations—smoking only and all three covariates—for both stomach and bladder cancer. For both sites of cancer, the distribution corresponding to the worst-case maximum discrepancy ratio was identical to that corresponding to the ratio for all three covariates. For bladder cancer, only the category "teachers," with a maximum confounding bias factor of 1.26, had a value above 1.20; 20 of 25 were below 1.10. For stomach cancer, there was even less effect: none of the confounding bias factors was above 1.15.

Discussion

Before discussing the substance of the findings, it is necessary to address some aspects of our approach. Confounding bias can be estimated accurately only if we know which covariates are true confounders and what is the optimal parametrization of these confounders and if we have very precise estimates of the true and of the inadequately adjusted odds ratio. Our estimates of confounding bias are undoubtedly exaggerated because the variation in odds ratio estimates embodies not only confounding bias but also statistical sampling error. Thus the various distributions of confounding bias summarized in Tables 4 to 7 would be closer to 1.00 if there were no sampling error. Furthermore, by presenting the maximum ratio of adjusted to unadjusted odds ratios, we highlighted the worst-case scenario.

Even if confounder variables have been identified and measured in a study and statistical adjustments have been carried out, this does not imply that the confounding has been completely controlled. On the contrary, residual confounding can arise in various ways: error in the measurement of the confounder covariate(s) and less than optimal stratification or modelling of the covariate(s). This is as much true in our attempts to control for confounding as it is for other investigators. However, we measured variables and conducted the statistical analysis in ways that are typical. The comparisons among different estimates and the inferences to be drawn are, therefore, likely to be representative of what other investigators may expect.

TABLE 7

Distribution of Confounding Bias Factors* for the Stratification by Smoking, Socioeconomic Status (SES), and Ethnicity Compared with the Estimate Unadjusted for These Variables, for the Associations Between Bladder Cancer and 25 Occupations and for the Associations Between Stomach Cancer and 25 Occupations

Confounding Bias Factor	Bladder Cancer		Stomach Cancer	
	Confounder Stratification†		Confounder Stratification†	
	Smoking	Smoking/ Ethnicity/ SES	Smoking	Smoking/ Ethnicity/ SES
1.00–1.10	23	20	24	20
1.10–1.20	1	4	1	5
1.20–1.30	1	1	0	0
Total associations	**25**	**25**	**25**	**25**

* For each cancer-occupation association, the confounding factor equals the greater of OR_i/OR_0 or OR_0/OR_i, wher OR_0 is unadjusted, and OR_i is adjusted for smoking only or for smoking, ethnicity and SES.

† All ORs were adjusted for age as well as for the variables indicated here.

The use of cancer controls is a common, albeit controversial, strategy in cancer case-control studies. However, it is not at issue here since our present purpose is not to derive valid odds ratios but rather to describe the variation in estimates generated by different strategies for the control of confounding. It is noteworthy, however, that, despite the use of cancer controls, the relative risk estimates for smoking vis-à-vis stomach, bladder, and lung cancer (shown in Table 2) are very similar to those found by Doll and Peto[17] in a cohort study among British physicians.

Finally, it should be emphasized that the inferences from the present study are by no means limited to case-control studies. That is, although the data bank and methods of analysis pertained to a case-control situation, the inferences about confounding bias are equally applicable to any epidemiologic design to estimate cancer-occupation relative risks.

One of the criteria used by epidemiologists to distinguish true from false associations is the strength of the association. That is, among two relative risk estimates which have equal levels of statistical significance but one of which is much greater than 1.0 while the other is closer to 1.0, the larger one is considered more likely to reflect a true association than the smaller one. This consideration follows from the recognition that some degree of bias is quite likely in any nonexperimental study. Small excess relative risks, even if they are statistically significant, are often interpreted with great caution, if not skepticism. Although there has been no explicit consensus on what level of excess relative risk should be considered too small to be taken seriously, we believe that many epidemiologists use a cut point in the range of 1.2 to 1.5 for this purpose.[18–20] Our results indicate that a cut point in this range is reasonable for studies of cancer-occupation associations. Of the 75 associations studied (three cancers, 25 occupations), only four manifested confounding bias factors in excess of 1.30. Of these four, only one exceeded 1.40, that for the lung cancer-teacher association. Thus the most extreme distortion was by a factor well under 2.00. For a study in which interoccupation group variation in smoking, socioeconomic status, and ethnic group is similar to that in our base population, a cancer-occupation relative risk estimate in the range of 2.00 or greater is most unlikely to be an artifact due to lack of adjustment by smoking, ethnic group, or socioeconomic status. Since these are among the major social factors commonly considered as confounders, it is unlikely that such associations would be due to confounding by other social

factors (excepting age, sex, and race which can usually be taken into account). In fact, our results would indicate that relative risks greater than 1.40 for lung cancer associations and greater than 1.20 for bladder and stomach cancer associations are unlikely to be attributable to confounding bias. On the other hand, our results also imply that relative risk estimates as low as 1.20 for lung cancer associations or 1.10 for bladder or stomach cancer associations run a fair chance of being attributable to confounding bias, even if they are "statistically significant."

The generalizability of our findings must be qualified. The variation between occupations in smoking, ethnic, and social characteristics may differ in other countries from what it is in Canada. Furthermore, the fact that our entire study population was drawn from a single city may mean that the interoccupation variation in smoking and social class was smaller than it would be in a national population in which occupation may well be associated with region of residence which may itself be an important correlate for smoking behavior. If this were the case, then our bias estimates would be lower than those pertaining to a national study, such as that reported by the Registrar General.[7] A second caveat concerns the possibility that workers in a single plant, which is usually the basis of a cohort study, may be more discrepant from the national norm in smoking and social class than are all members of a particular occupation group. Our study concerned all members of the 25 selected occupation groups who entered the study sample. Typically, the subjects in each occupation group were drawn from a wide variety of industries and workplaces. A given company or plant may have had some recruitment practices that would have had the effect of selecting a work force with ethnic and smoking characteristics quite atypical of that occupation as a whole. In a cohort study of such a work force, confounding bias could be greater than in our study.

Given the strength of the association between smoking and lung cancer, it was not surprising that the confounding effects, such as they were, were more pronounced when smoking was adjusted for than when SES or ethnicity were adjusted for. The joint confounding effect of multiple covariates can, in theory, be much more serious than the effect of each individually.[4] It is thus of interest to note that we found little in the way of a joint confounding effect. Adjusting for all three covariates had slightly more effect than adjusting for smoking alone.

In the comparison of estimates based on different stratifications of smoking, it was noteworthy that the three category variables based on duration and pack-years performed much better than the simple two-category smoking variable "never/ever." Five-category smoking variables performed only slightly better than the three-category variables. The present analysis was not a systematic comparison of confounding resulting from alternative smoking stratifications. It may be that different cut points from the ones used on the smoking intensity scale would have provided effects equivalent to those seen for the duration and pack-years variables; however, we doubt it. The cut points on the three scales provided roughly equal distribution of subjects across the various strata, and the odds ratio estimates compared with nonsmokers were about equal in the different strata for the three variables. The differences in effects of "intensity" as a confounder as opposed to "duration" and "pack-years" are likely due to lesser variability in intensity than in duration across the different occupations.

There have been two studies similar to this one. Asp,[10] using Finnish data on interoccupation variation and making some reasonable assumptions on lung cancer risks in different smoking categories, algebraically derived the confounding bias factors that could be expected in lung cancer-occupation associations that were not adjusted for smoking. The results were similar to ours in that 23 of 25 occupations groups manifested confounding bias factors less than 1.30 and none exceeded 1.50. Blair et al,[11] using data from a large American cohort study, carried out a fully empirical analysis, as we did, to estimate associations between three types of cancer (lung, bladder, intestine) and a large number of occupations, adjusted for smoking only and unadjusted. The only associations which manifested confounding bias factors above 1.30 were based on very small numbers, and statistical variability was as likely an explanation as bias. The results were in general similar to ours, although theirs were expressed in terms of correlation coefficients between adjusted and unadjusted values. The similarity of findings in Finland, the United States, and Montreal lends some weight to their generalizability.

The optimal strategy in any study is to collect as much quality information on potential confounders as possible as well as on the disease and exposure of interest. Not only does this permit some means of reducing or eliminating confounding bias but it also provides a means of detecting effect modification. However, given limited resources for epidemiologic investigation, and limitations inherent in data sources, it may be necessary to compromise on data quality and quantity. In setting priorities, investigators should concentrate on the "first-order" elements of a study, namely, appropriate selection of study subjects and reliable measurement of the disease and exposures of interest. Confounding is a "second-order" element of a study and should generally be accorded the attention that this ordering of priorities implies.

The epidemiologic literature on occupational cancer is dominated by cohort and national death certificate standardized or proportional mortality ratio studies in which there has been no control for confounders other than age, sex, race, and possibly state or county of residence. In the absence of evidence about the strength of the possible confounders, it is legitimate to question how much weight to attribute to this body of literature. Our results, plus those of others,[10,11] lead to the encouraging conclusion that findings from such studies, where smoking, SES, and ethnicity were not available, are unlikely to be seriously distorted by confounding due to these factors. Our findings also imply that the inability

to collect information on smoking or another potential confounder should not, in itself, be considered a fatal flaw in any proposed study, although prudence would usually dictate that it should be collected if technically and financially feasible.

We certainly do not advocate uncritical acceptance of cancer-occupation results or any others. Confounding bias can, in some circumstances, cause serious distortions and one should be concerned to avoid confounding bias if one is planning a study or to evaluate its impact if one is interpreting a study. However, often there is little information to go by in trying to estimate the impact of missing confounder information. Our findings may provide a guidepost. Of course, confounding is one of several sources of bias in a typical epidemiologic study, and all sources should be considered when interpreting results.

Acknowledgments

The project from which these results were derived was funded by National Health Research and Development Program of Canada, the National Cancer Institute of Canada, and the Institut de Recherche en Santé et Securité du Travail du Québec, who also supported Ms Greenwood on a student grant.

Coding of jobs was done by Howard Kemper. Case ascertainment and interviewing were done by Denise Bourbonnais, Yves Céré, Lucy Felicissimo, Hélène Sheppard, Vincent Varacalli, and Michel Vinet. Jean Pellerin helped with data management. This study would not have been possible without the cooperation of the following clinicians and pathologists: Drs Y. Méthot, R. Vauclair, and Y. Ayoub, Hôpital Notre-Dame; Dr R. Hand, Royal Victoria Hospital; Drs C. Lachance and H. Frank, Sir Mortimer B. Davis Jewish General Hospital; Drs W.P. Duguid and J. MacFarlane, Montreal General Hospital; Drs S. Tange and D. Munro, Montreal Chest Hospital; Drs F. Gomes and F. Wiegand, Queen Elizabeth Hospital; Drs B. Artenian and G. Pearl, Reddy Memorial Hospital; Drs D. Kahn and C. Pick, St Mary's Hospital; Dr C. Piché, Hôpital Ste-Jeanne d'Arc; Drs P. Bluteau and G. Arjane, Centre Hospitalier de Verdun; Drs Y. McKay, and A. Bachand, Hôpital du Sacré Coeur; Drs A. Neaga and A. Reeves, Hôpital Jean-Talon; Drs Y. Boivin and M. Cadotte, Hôtel-Dieu de Montréal; Dr A. Iorizzo, Hôpital Santa Cabrini; Dr A. Bonin, Hôpital Fleury; Drs J. Lamarche and G. Lachance, Hôpital Maisonneuve-Rosemont; Drs G. Gariépy and S. Legault-Poisson, Hôpital St-Luc; Dr M. Mandavia, Lakeshore General Hospital; Dr J.C. Larose, Cité de la Santé. The authors thank the pathology department and tumor registry staff of the above-mentioned hospitals who notified them of new cases. The manuscript was prepared by Ms C. Dupuis and Ms L. Carfagnini.

References

1. Bross JJ: Pertinency of an extraneous variable. *J Chronic Dis* 1967, 20:487–495.
2. Miettinen OS: Components of the crude risk ratio. *Am J Epidemiol* 1972;96:168–172.
3. Schlesselman JJ: Assessing effects of confounding variables. *Am J Epidemiol* 1978;108:3–8.
4. Breslow NE, Day NE: *Statistical Methods in Cancer Research*, vol I: *The analyses of case-control studies*. IARC Scientific Publications No. 32, Lyon International Agency for Research on Cancer, 1980.
5. Greenland S, Robins JM: Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;15:413–419.
6. Thomas DC, Greenland S: The efficiency of matching in case-control studies of risk-factor interactions. *J Chronic Dis* 1985;38:569–574.
7. The Registrar General's Decennial Supplement for England and Wales 1970–1972: Occupational mortality series DS No. 1. London: Her Majesty's Statistics Office, 1978.
8. Cornfield J, Haensel W, Hammond EC, et al: Smoking and lung cancer: Recent evidence and a discussion of some questions. *JNCI* 1959;22:173–203.
9. Yanagawa T: Case-control studies: Assessing the effect of a confounding factor. *Biometrika* 1984;71:191–194.
10. Asp A: Confounding by variable smoking habits in different occupational groups. *Scand J Work Environ Health* 1982;10:325–326.
11. Blair A, Hoar S, Walrath J: Comparison of crude and smoking-adjusted standardized mortality ratios. *J Occup Med* 1985;27:881–884.
12. Siemiatycki J, Day N, Fabry J, et al: Discovering carcinogens in the occupational environment: A novel epidemiologic approach. *JNCI* 1981;66:217–225.
13. Siemiatycki J, Gérin M, Richardson L, et al: Preliminary report of an exposure-based, case-control monitoring system for discovering occupational carcinogens. *Teratogenesis Carcinog Mutagen* 1982;2:169–177.
14. Siemiatycki J, Richardson L, Gérin M, et al: Associations between several sites of cancer and nine organic dusts: Results from an hypothesis-generating case-control study in Montreal, 1979–1983. *Am J Epidemiol* 1986;123:235–248.
15. Canadian Classification and Dictionary of Occupations, vol 1. Dept of Manpower and Immigration, Immigration Canada, Ottawa, Canada 1971.
16. Dewar R, Siemiatycki J: A program for point and interval calculation of odds ratios and attributable risks from unmatched case-control data. *Int J Biomed Comput* 1985;16:183–190.
17. Doll R, Peto R: Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J* 1976;2:1525–1536.
18. Day NE: Epidemiological methods for the assessment of human cancer risk, in Clayson DB, Krewski D, Munro I (eds): *Toxicological Risk Assessment*, vol 2, CRC Press, Boca Raton, FL, 1985.
19. Monson R: *Occupational Epidemiology*, CRC Press, Boca Raton, FL, 1980.
20. Robins JM, Landrigan PJ, Robins TF, et al: Decision making under uncertainty in the setting of environmental health regulations. *Public Health Policy* 1985;6:322–328.